

早稲田大学大学院理工学研究科

# 博士論文審査報告書

論 文 題 目

パラ言語情報を利用した

音声対話システムに関する研究

Studies on Spoken Dialogue System

Using Para-Linguistic Information

申 請 者

藤江 真也

Shinya Fujie

電気工学専攻・知覚情報システム研究

2 0 0 5 年 3 月

人間と自然に対話できる機械システムを実現することは、人工知能研究が目標とする象徴的な課題のひとつである。本論文では、この自然な対話の実現という問題に対し、パラ言語理解能力の付与という観点から取り組んでいる。

従来の音声対話研究は、言語情報の扱いに偏重して進められてきた。ここで言語情報とは、発話を文字に書き起こしたときそこに現れる情報をいう。これに対し、人間同士の自然な音声対話では、文字に書き起こすことができない情報が重要な役割を担っていることが少なくない。例えば、話者の心情が言語の形をとらず発話の韻律や顔の表情に表れ、これを対話相手がくみ取ることで自然なやりとりが実現されることがよくある。また、対話相手の発話の調子にあわせて相槌や復唱を行うことにより、リズムある対話を実現する場合もある。このとき韻律や表情に表れた話者の心情や、相槌・復唱が担う対話調整の情報のように、発話に付随して生成され、言語情報の円滑な伝達を補助する非言語情報をパラ言語情報と呼ぶ。言語情報が、意識のレベルの高いところで（強く意識をして）授受されるのに対し、パラ言語情報は意識のレベルの低いところで（あまり意識することなく）授受される。このため、通常その存在を意識することは少ないが、これがないと対話は非常に不自然なものとなる。本研究では、このパラ言語の理解に焦点をあて、その抽出手法とそれを利用した対話システムの構成について検討を行ったものであり、6章から構成される。以下に各章の概要を述べるとともに、その評価を述べる。

第1章では、序論として音声対話において対話者間で交換される情報の種類を分類し、その中で特にパラ言語の担う役割を整理し、対話においてパラ言語に注目することの重要性について述べている。

第2章では、パラ言語情報の抽出において重要な鍵と成る、音声の基本周波数の抽出法について概観している。基本周波数は、人によって知覚される音声のピッチと高い相関を持つ情報で、発話の調子を表す量として有用な情報である。本研究のような韻律に着目した研究が可能になったのは、基本周波数の推定を実時間で精度良く行える技術が進んだことに負う部分が多い。ここでは、音声信号の相互相関を用いる手法、動的計画法を用いる手法、瞬時周波数を用いる手法、楕型フィルタを用いる手法をとりあげ、これらについて詳説している。

第3章では、本研究で対象とする対話タスクと、プラットフォームとして用いる対話ロボットの概要が述べられている。従来多くの音声対話研究が対象としてきた対話を、検索型対話として分類した上で、本論文で扱う対話を意思決定支援型対話と定義している。検索型対話では、ユーザが要求の具体案を事前に明確に持つことを前提とし、実際の対話ではユーザはその要求をシステムに対し明示的に伝え、システムはその条件を満たす案件を検索してユーザに提示するといった対話であるのに対し、意思決定支援型対話では、ユーザには事前に

要求の具体像はなく、対話を通じてこれを明確化することが対話の目標となる。双方の比較をすることで、意思決定支援型対話が、より高度な意図理解の問題を含み、よりパラ言語情報の扱いを必要としていることを述べている。

第4章では、意思決定支援型対話における、システムの提案発話に対するユーザの応答発話の意図理解に焦点を当て、そこにおけるパラ言語情報の利用について検討している。このような場面における、ユーザの応答発話は、言語的には提案を肯定的に捉えたか否かを判断できない曖昧な表現であることが多いが、韻律や表情・頭部動作などから、話者の心情（提案を肯定的に受け取ったか否か）を理解できる。本研究では、まず発話の韻律情報を用いて、発話が提案に対し、肯定的か否定的かを認識する手法を提案している。韻律のダイナミックレンジや最終モーラ長などが、この種の発話の分類に重要であることを示し、最終的に人と同程度の精度で発話を分類することに成功している。ついで、肯定を表す頭部ジェスチャとしてうなずきを、否定を表す頭部ジェスチャとして、かしげ、首振りを取り上げ、これらの認識方式を提案している。頭部領域のオプティカルフローを特徴量として用い、隠れマルコフモデルにより認識を行うことで、自然な対話中に含まれる非常に微妙な頭部動作を捉えることができることを示している。また、これらの認識機能を持つ対話システムを実装し、意図理解にパラ言語情報を用いることの有効性を確認している。ここで実現された対話システムでは、言語以外の情報から話者の意図を推定できることで、発話のやりとりを簡素化することができ、従来にない効率的なリズムのある対話を実現できており、高く評価できる。

第5章では、システムの発話タイミング制御におけるパラ言語の利用法について、特にユーザの発話に対する相槌や復唱の生成の観点から検討している。相槌や復唱は、対話相手に円滑に発話を促すため重要な役割を演じるが、どのタイミングでどのような内容の発話を挟むかが重要な意味を持つ。一般に人同士の対話では、相槌・復唱は相手の発話中に頻繁に挟まれることになるが、従来の音声認識では相手の発話終了を待たないとその内容を認識できないため、このような相手の発話中での、相手の発話内容に応じた相槌・復唱の挿入は困難であった。また、相槌・復唱を入れるタイミングについても検討は不十分であった。この問題の解決のために、本研究では、まずユーザの発話終了を待たずに認識結果を早期確定する音声認識器を、有限状態トランスデューサ（FST）を用いて構成している。FSTを用いてユーザの発話内容とこれに対する適切な相槌・復唱内容との関係を記述しておくことで、どういう条件でどのような内容を返すべきかをいち早く準備できる枠組みになっている。つぎに、相手発話の韻律パターンから、相槌・復唱が許されるタイミングを整理し、FSTで用意された発話を適切なタイミングで出力する枠組みを実現している。これらによって、適切なタイミングで、適切な内容のフィードバックを行うことが可能になっている。本手法は、ユーザの発話内容を予め静的に記述する必要があると

いう制約はあるが、発話内容を発話終了を待たずして早期に確定できる従来にない画期的な枠組みであり、これが基礎となって、自然なタイミングでの応答が実現されている。この技術の応用の可能性は広く、成果は高く評価できる。

第6章は、結論であって、論文のまとめと今後の課題が述べられている。

以上を要するに、本研究では、従来の音声対話システムにおいては考慮されることの少なかったパラ言語情報に着目し、対話進行に有用なパラ情報を抽出する様々な手法を開発するとともに、適切な内容のパラ言語情報を適切なタイミング生成するための機能について検討し、従来にないリズムある対話を可能とするシステムを実現したものであって、その工学的価値は高い。よって、本論文は、博士（工学）の学位にふさわしいものと認める。

2005年2月

#### 審査員

(主査)	早稲田大学教授	工学博士（早稲田大学）	小林 哲則
	早稲田大学教授	工学博士（早稲田大学）	成田誠之助
	早稲田大学教授	工学博士（早稲田大学）	橋本 周司
	早稲田大学教授	工学博士（早稲田大学）	
		Ph.D(Stanford 大学)	松山 泰男
	早稲田大学教授	工学博士（早稲田大学）	笠原 博徳